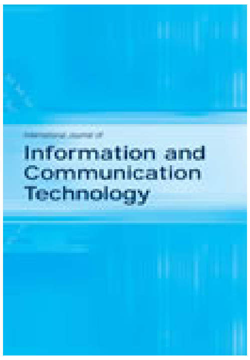


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Deep learning in machine translation: revolutionising language processing

Wei Xiu, Limei Gong, Zhifei Wang, Jianing Liu

DOI: 10.1504/IJICT.2025.10071312

Article History:

Received:	17 March 2025
Last revised:	16 April 2025
Accepted:	16 April 2025
Published online:	27 May 2025

Deep learning in machine translation: revolutionising language processing

Wei Xiu

Mathematics and Computer Department,
Chaoyang Normal University,
Liaoning, Chaoyang, 122000, China
Email: xiuwei@cynu.edu.cn

Limei Gong

Foreign Languages Department,
Chaoyang Normal University,
Liaoning, Chaoyang, 122000, China
Email: gonglimei@cynu.edu.cn

Zhifei Wang*

Mathematics and Computer Department,
Chaoyang Normal University,
Liaoning, Chaoyang, 122000, China
Email: 15942145577@163.com
*Corresponding author

Jianing Liu

Scientific Research Office,
Chaoyang Normal University,
Liaoning, Chaoyang, 122000, China
Email: 15124083555@163.com

Abstract: Machine translation (MT) has undergone a remarkable transformation with the rise of deep learning methods, significantly improving translation accuracy and fluency. This paper examines how deep learning methodologies have influenced MT, particularly through the use of neural networks, Seq2Seq models, attention mechanisms, and transformer architectures. Traditional rule-based or statistical approaches have evolved into neural machine translation (NMT), leveraging large-scale data and advanced learning paradigms. The study highlights advancements, challenges, and future prospects, focusing on low-resource language translation, model bias, and computational efficiency. By analysing current developments and trends, this paper emphasises the revolutionary role of deep learning in enhancing multilingual communication through machine translation.

Keywords: deep learning; machine translation; MT; neural networks; transformer models; language processing.

Reference to this paper should be made as follows: Xiu, W., Gong, L., Wang, Z. and Liu, J. (2025) 'Deep learning in machine translation: revolutionising language processing', *Int. J. Information and Communication Technology*, Vol. 26, No. 16, pp.1–15.

Biographical notes: Wei Xiu is a research scholar in the Departments of Mathematics and Computer Science at Chaoyang Normal University, which is in Liaoning, China. His main academic research areas are mathematics and computer science and he is known for his work in both the theoretical and practical areas of the field. After completing her studies in foreign languages and linguistics,

Limei Gong has been actively engaged in teaching and research at Chaoyang Normal University in Liaoning, China. With a strong academic foundation in applied language studies, she has contributed to curriculum development and scholarly research within the Foreign Languages Department. Her recent work explores innovations in language instruction and intercultural communication.

Zhifei Wang is a research scholar at the Mathematics and Computer Department, Chaoyang Normal University, in Liaoning, China. With a focus on mathematics and computer science, Wang contributes to academic research and technological development in these fields.

Jianing Liu is a researcher at the Scientific Research Office, Chaoyang Normal University, in Liaoning, China. She is actively involved in academic development and contributes to the advancement of research initiatives within the university.

1 Introduction

Machine translation (MT) has been significantly improved by recent advances in deep learning and artificial intelligence (AI). Indeed, the advancement of AI and deep learning into the mechanics of language has made it possible for the creation of translation systems capable only of producing translations that are very accurate and fluent (Mi and Xie, 2024). Deep learning, a part of machine learning that is based on artificial neural networks which are combined into multiple layers, has noticeably achieved better results than traditional statistical and rule-based translation methods (Chandrakala et al., 2024). The change from the phrase-based kind of the SMT method to the neural machine translation (NMT) kind of translation has been a proper way to improve the quality of translation to the level that it can be integrated into and can be reliable for, practically all applications (Oliveira et al., 2024).

Even in the increasingly globalised world, the role of MT is really significant, and there are no two opinions about that. It is used for cross-cultural communication, it supports the global process of business, and it backs up international cooperation (Guimarães et al., 2024). From the translation of internet content to the conversion of real-time speech to text, MT systems are very important tools for individuals and companies as well (Shenbagaraj and Iyer, 2024). However, the complex systems of human language, such as syntax, semantics, and context, still are the main obstacles standing before the translation systems. Classical MT systems had guides with rules developed by linguists or models of probabilistic scenarios, neither of which could reflect linguistic effects (Ahda et al., 2024). The acceptance of the deep learning method has led to a situation whereby a number of challenges can find their answers in the development of a data-driven and, thus, the more effective language representation learning model, similar to human beings.

The recent employment of deep learning models in many language tasks reveals a situational approach (NLP) for example text classification detection, sentiment analysis, and speech recognition, the success of which is remarkable. It is a dramatic leap from conventional methods for language pattern recognition to the introduction of sequence-to-sequence or attention-based (Seq2Seq) models, which are popularly known in the field of MT (Sohrab et al., 2024). One cannot overlook that the mechanism enables attention models to more precisely process input sequences and create output sequences that blend better with the original source. One essential factor has led to the liberated sequential use of recurrent dynamics and long short-term memory (LSTM) devices in the MT realm which is the advent of the transformer architecture (Kaur and Chauhan, 2024). Through the self-attention mechanism and the positional encoding of the input feature vector, transformer models were able to meet a much greater quality translation rate than any previous systems thus becoming a kind of symbolic backbone of the most advanced translation systems (Jing, 2024).

Deep learning technologies in the field of MT have the benefit of being soundly adapted to various languages and domains, which is one of their strengths. In contrast to classical methods that focused on heavy manual feature engineering, deep learning models inferred features directly from huge non-structured corpora. OpenAI's GPT and Google's BERT are both examples of models that can be considered features that paved the way for language processing systems that have a multilingual and contextual flavour (An et al., 2024). Additionally, drawing on the knowledge obtained from richer languages through the transfer of learning methods helped boost the translation of resource-poor languages (Shang and Li, 2024). Although the translation industry has seen large advances in this area, there are still major limitations such as lack of data about the target language, the extensive computations necessary for any form of automation, and biases being introduced by skewed training data. Efforts to solve these challenges and improve the quality of MT systems will contribute greatly to guaranteeing the moral, fair, and sustainable global use of this technology (Zemni et al., 2024).

Another critical aspect of deep learning in MT is its ability to capture contextual meaning and idiomatic expressions. Whereas idioms, colloquialisms, and ambiguous phrases frequently cause difficulties in traditional translation systems, the modern system is much more reliable. The reason is the incorporation of deep learning models such as attention and transformers (Zemni et al., 2024). Human-like translations that are fluent and accurate are produced by these systems thanks to considering much broader context than originally the case. In addition, humanity continues to respond to the surprising quality of machines. As a result, human feedback on the rendered translations has become a method to improve these models by reinforcement learning (Abdul Rauf and Yvon, 2024).

Incorporating deep learning into MT not only involves translating text but also requires the handling of texts utilising different languages. The newest technologies made it possible to translate between the languages of the text and, for example, pictures, sounds, or videos. This technology potential is significantly changing sign language protocols as well as speech inputs along with integrative interactions with devices like virtual assistants, among other things (Ju et al., 2024). Major companies like Google, Microsoft, and Meta are already rising this trend. They have neutral translations but using language in different styles according to what the user's purpose is and the hardware they are using. The development of deep learning applications in mobile (Wei et al., 2024) and cloud platforms has further contributed to MT efforts, ensuring that together people from different language backgrounds can interact seamlessly.

Though some impressive advances have been made, serious issues still exist regarding the creation of robust and unbiased text translations systems. A primary source of this problem is unbalanced training databases from which MT algorithms are learned that operate purely on text data. Biases in the input data inevitably translate into biased outputs which may lead individuals to hold wrong beliefs or misconceptions due to such a translation (Karabayeva and Kalizhanova, 2024). Techniques such as adversarial training which are designed specifically for protecting the output from being negatively influenced by bias present in data during model training offer the hope of alleviating this problem (Anderson et al., 2024). It indeed includes a plethora of solutions but one still remains very urgent to tackle – the excessive energy use and accessible nature of training such heavy deep learning models. It is thus within the boundaries of this problem that research activity will mainly focus on converting the models into very effective ones.

Furthermore, while deep learning has been a great asset in not only reaching human-level translation but improving translation quality, it is a complicated goal. The most vulnerable languages still cannot be translated optimally, especially those with scanty digital resources (Kjell et al., 2024). The research in semi-supervised and unsupervised language learning has yielded the hope of improved low-resource language translation, thus utilising unlabeled data and cross-lingual embeddings. Deep learning evolves and improves, and thus it is possible to create translation systems that are more diverse and equal, enabling a more comprehensive linguistic representation and access. The objectives of this study are:

- To analyse the impact of deep learning on MT, explaining specific advancements such as attention mechanisms and transformer architectures.
- To identify the challenges and forecast the future of deep learning in MT, which includes issues about bias, computational efficiency, and the translation of low-resource languages.

To sum up, deep learning has given a new face to the industry of MT, making the work not only correctly done, easy, and in the right language but also accurate, fluent, and adapted to the context. The transformation from old-fashioned ways to ones based on the neural networks has made a tremendous step that the involved translation systems have become smarter and could be easily used by the wider masses. Nevertheless, the fact is that problems such as bias, constraints due to computation, and the under-representation of low-resource languages are still issues for which more focus on research and innovation is needed (Zhang et al., 2024). In the future, deep learning can be an integrated part of other AI-driven technologies which will make multilingual communication and cross-cultural understanding better. The research gives a broad understanding of the influence of deep learning on MT, highlighting the power of the technology to improve it while also pointing out the obstacles ahead.

2 Literature review

The integration of deep learning into MT has been a hot topic for the past several years, and a lot of research has been done to study innovative neural architecture, experimental training techniques, and various measures aimed to improve translation quality. Very much so that this shift from statistical machine translation (SMT) to NMT has been revolutionary in terms of accurate standards, smoothness of delivery, and sense of context; yet this has also affected the nature of the studies and the changes in methodologies applied in these areas that have had a huge effect on the translations done by these systems (Patle and Kalra, 2024). This paper is going to present an overview of important work done in this area along with the main conclusions arrived at, methods used, as well as comments left to be considered for subsequent improvements in the field. In particular, the authors refer to confusing hence submissions that looked into how further developments of sequence-to-sequence models have augmented interfaces that connect prior and subsequent words while also giving credit to transformer-based structures and their ability to identify which words in one language stand for which in the other without first explaining them beforehand and thus breaking down the sentence further until it is understandable to people who speak those languages originally. The analysis will not only identify areas for growing possible improvements but also ways to steer towards the developments which will potentially take dead technology even further than it.

The study by Luo et al. (2024) concentrates on the interpretability of deep learning models and their application in natural language processing (NLP) while dealing with the problems brought by opaque models. Their work classifies interpretability methods into three major categories: finding the relations between input features, being aware of the explained feature and giving explanations by natural languages, and activating the hidden nodes in the neural networks. Through a complete and systematic approach to the definition and components of interpretability, the outcome of this study is sure to add to the clarity of NLP technologies such as MT and sentiment analysis through actively engaging users.

In their study, Nam and Jang (2024) probed the emerging technology of multimodal machine learning by emphasising the cyclical transformation between images and textual representations. The research team went through the current metamorphoses concerning the introduction of deep learning various architectures, and the linkage of both modalities citing real-life applications' dependence on multimodal learning were the developers of such innovations. The contribution of the study consists of a collection of the most recent development paradigms, general performance measure analyses, and hints on follow-up studies to enhance the study of language mapping onto graphics and vice versa.

Belete et al. (2024) put forth a word sense disambiguation model for the Ge'ez language, a mission highly complicated due to the lack of public datasets and WordNet resources. They amassed a collection of 1010 ambiguous word samples, through which they inferred the meaning from the context using a variety of modern text vectorisation facilities including bag of words, TF-IDF, and word embeddings. Their analysis signified that random forests together with the bag of words gave 99.52% accuracy whilst deep learning technologies such as LSTMs and DNNs provided 100%. They demonstrated the superiority of deep learning by successfully addressing the linguistic ambiguity problem.

Macas et al. (2024) conducted a study on the various ways that adversarial attacks could be directed at deep learning-based cybersecurity systems. The classification of attack types is based on, for instance, the location where it would occur, the objectives and technologies employed by the attackers themselves, and, finally, the defensive strategies currently existing. The research demonstrates the inadequacies of deep learning models employed in cybersecurity tasks, and thus possible countermeasures were assessed to eliminate risks. This research is deemed to be a significant starting point in the field of adversarial attacks which should help to enhance the safety of AI-supported cybersecurity systems.

Han et al. (2024) evaluated the efficacy of transferring semantic and knowledge information to improve the clinical text translation process in English. For this purpose, they used a Transformer network based on a multilingual neural system. Supported by the learned capabilities of extraction, as well as the appropriate preliminary training of MMPLMs, their models achieved a remarkable performance of around 80% in English-Spanish clinical domain data at the ClinSpEn-2022 event organised as a follow-up of the English-Spanish bilingual course of the conference. Notably, even smaller pre-trained language models were shown to have an edge over bulkier ones in the task of the specific field fine-tuning. This finding is significant and relevant for the respective research activities in clinical MT and health text analytics.

Sunar and Khalid (2024) investigated how NLP technologies can help in improving students' open-ended replies in the classroom environment. They synthesised twenty-eight articles up-to-date in the field of NLP to extract their insights which were then grouped into six themes, including but not limited to, methodologies, models, and data characteristics patterns. Their research provides an organised comparison of the data, which is a valuable source of information for those who are looking to use NLP in analysing educational feedback and teaching issues.

Mahdi et al. (2024), have come forward with great efforts in the area of understanding the part of deep learning in handwritten character recognition, especially in the case of Arabic and Urdu as they have been the least focused on in research. The study shows that the difficulty of lucid handwriting comes from different people using different writing styles and different types of writing instruments. They have compared the various deep learning techniques used in optical character recognition and the results of the study emphasise that deep learning is a superior solution over traditional machine learning for the case of pattern recognition and computer vision applications.

In their study, Pradhan and Yajnik (2024) investigated the task of parts-of-speech (POS) tagging for the low-resourced language Nepali. The research compared hidden Markov models (HMM), the conditional random fields (CRF), and LSTM networks, and the conclusion was that LSTM has surpassed the other competitors with an accuracy figure of 99.6%. Their work has shown how powerful deep learning can be in dealing with complex morphological structures in Nepali and as a consequence, it has led to much better NLP programs for languages with low resources.

Ni et al. (2024) started a new study with the introduction of the Text-to-GraphQL (Text2GraphQL) task, which is used to change users' questions into GraphQL statements for the graph database. This work is different from the traditional ones, Text-to-SQL tasks, which are normally based on SQL, whereas this research brought about a pipeline which is a mixture of a pre-trained adapter and the pointer network to enforce semantic parsing. Their method facilitates human-machine communication by the efficient conversion of natural language queries into structured database retrievals, especially for the medical sector human-robot interaction (HRI) applications where it serves the most significance.

Table 1 Literature comparison

Author(s)	*Journal*	*Focus area*	*Methodology*	*Key findings*
Luo et al.	*ACM Computing Surveys*	Interpretability in NLP deep learning models	Local interpretation through input features, natural language explanation, and probing hidden states	Improved model transparency
Nam et al.	*Expert Systems with Applications*	Multimodal ML for bidirectional translation	Integration of NLP and image processing	Developed taxonomy for multimodal learning
Belete et al.	*Ampersand*	Word sense disambiguation for Ge'ez language	ML-based techniques like Random Forests and Deep Learning	Achieved 100% accuracy using deep learning
Macas et al.	*Expert Systems with Applications*	Adversarial attacks on ML in cybersecurity	Categorisation of attacks and defense mechanisms	Identified key trends in adversarial ML
Han et al.	*Frontiers in Digital Health*	Clinical text machine translation	Transformer-based models, transfer learning	Effective multilingual translation for clinical texts
Sunar et al.	*IEEE Transactions on Learning Technologies*	NLP for student feedback analysis	Sentiment analysis, topic categorisation	Comprehensive synthesis of student feedback studies
Mahdi et al.	*Multicriteria Algorithms with Applications*	Handwritten character recognition in Arabic and Urdu	Deep learning-based OCR models	High accuracy in Arabic-adapted script recognition
Pradhan et al.	*Multimedia Tools and Applications*	POS tagging for Nepali language	HMM, CRF, LSTM	LSTM achieved highest accuracy of 99.6%
Computer Science et al.	-	NLP research overview	Various NLP applications and deep learning advances	Highlighted NLP impact across industries
Ni et al.	*Information Systems Frontiers*	Text-to-GraphQL conversion for databases	Pre-trained adapter, pointer network	Effective Text2GraphQL task implementation

3 Methodology

The research methodology for deep learning in MT embraces a multi-pronged strategy that includes data gathering, selection of models, training procedures, evaluation metrics, and finally a model that has been proposed that is intended to improve translation quality. In this study, both theoretical and experimental analyses were conducted with the aim of gaining a profound understanding of how deep learning has transformed language processing in MT. The methodical outline of this section describes the distinct steps which one has to go through while developing, testing, and validating NMT systems.

3.1 Data collection and pre-processing

Deep Learning-based MT training can be successful only when the high-quality parallel corpus is available for the translation. The data we adopted came mainly from the Europarl Corpus, the workshop on machine translation (WMT) datasets, and the OpenSubtitles datasets. The above-mentioned datasets contain translations that were aligned at the sentence level for several language pairs. These datasets provide us with the necessary textual resources for deep learning that can recognise linguistic structures, grammar, and context. Before proceeding further, the dataset must be processed so as to achieve greater efficiency and accuracy in the models trained. Pre-processing is the first step in this process, which is followed by such stages as tokenisation, which is separating the text into smaller entities like words or subwords. This procedure is meant to facilitate the processing of neural networks. Byte pair encoding (BPE), a subword tokenisation technique, is also employed to manage rare and out-of-vocabulary words. Additionally, the data normalisation techniques such as lowercasing, and punctuation elimination, as well as filtering out the noisy sentences, are used to ensure that there is consistency in the training data. The cleaned-up dataset is subsequently divided into training, validation, and test sets using which the performance of the model will be accurately measured.

3.2 Model selection and training procedure

Deep learning-based MT systems have undergone a major change from traditional sequence-to-sequence (Seq2Seq) architectures to attention-based transformer models. The transformer or the Transformer model introduced by Vaswani et al. in 2017 is now the base of the neural machine translating system thanks to its ability to process long-range dependencies in a very effective way. It is different from both recurrent neural networks (RNNs) and LSTM networks which operate sequentially. The Transformer, in contrast, processes the whole input sequence at once by self-attention which reduces training time and the possibility of fluency in translation.

In this research, the Transformer model has been adapted. The bidirectional encoder representations from transformers (BERT), generative pretrained transformer (GPT), and text-to-text transfer transformer (T5) pre-trained models have been utilised. These models have been shown to perform better than traditional ones in different tasks of NLP like translation by taking advantage of the well-understood prior knowledge from context. We use the Adam optimiser to dynamically adjust learning rates during training, improving convergence and translation quality. The training also uses the cross-entropy loss function as the objective to minimise the difference between the predicted translations and the actual ones to be translated.

To further improve the quality of the translations, transfer learning will be used as a technique in which the model is first trained on high-resource languages and then on low-resource languages to fine-tune the model. This technique has the potential of overcoming the problem of the unavailability of data and can therefore create solutions applicable across different linguistic systems. Also, the model is made more robust by means of data augmentation techniques such as back-translation (where target language sentences are translated back to the source language to generate synthetic training data).

3.3 Evaluation metrics and performance assessment

The assessment of MT models frequently couples automated metrics and human evaluations to determine translation accuracy and fluency. Here, we use bilingual evaluation understudy (BLEU), metric for the evaluation of translation with exact ordering (METEOR), and, as the main evaluation metric, the translation edit rate (TER).

- BLEU score: this determines how many terms or phrases in the translation resemble those in the original language. The higher the score is, the more accurate the translation is.
- METEOR score: METEOR is a more precise measure of fluency and semantic correctness than BLEU, as it accounts for synonyms, stemming, and word order.
- TER: the rate of translation requiring humans to make changes on the computer-translated output to ensure language compatibility is assessed here. A lower TER means a translation is more accurate. As an additional assurance of reliability, human evaluation contributes to the process where translations are assessed by bilingual experts based on their grammatical, coherent and cultural correctness. This stage is very vital to ensure the credibility of the results obtained from the deep learning model exceeds the numerical.

3.4 Challenges and solutions in neural MT

Though neural MT has made great strides, there are still some hurdles to be cleared. On the one hand, there exists the problem of computational cost since the performance of deep learning models often demands enormous resources for their training. This can be a problem for researchers who cannot access enough resources to train large models. The approach taken is `the right one that works to local-scale model optimisation with minimal computation which makes the use of techniques (e.g., model pruning and quantisation) possible. This means that the model density can be reduced, and its function seems still high.

Bias in training data is another major issue in language models, as they often learn stereotypes from large corpora and then propagate those stereotypes. This research employs bias-mitigation methods such as fairness-aware learning, dataset rebalance, and controversy-solving methods such as the use of adversarial training when developing MT systems.

Furthermore, the scarcity of bilingual data makes it very hard to make a good translation for low-resourced languages. To address this problem, unsupervised and semi-supervised learning techniques are studied where the models follow some self-supervised targets for learning at a monolingual level.

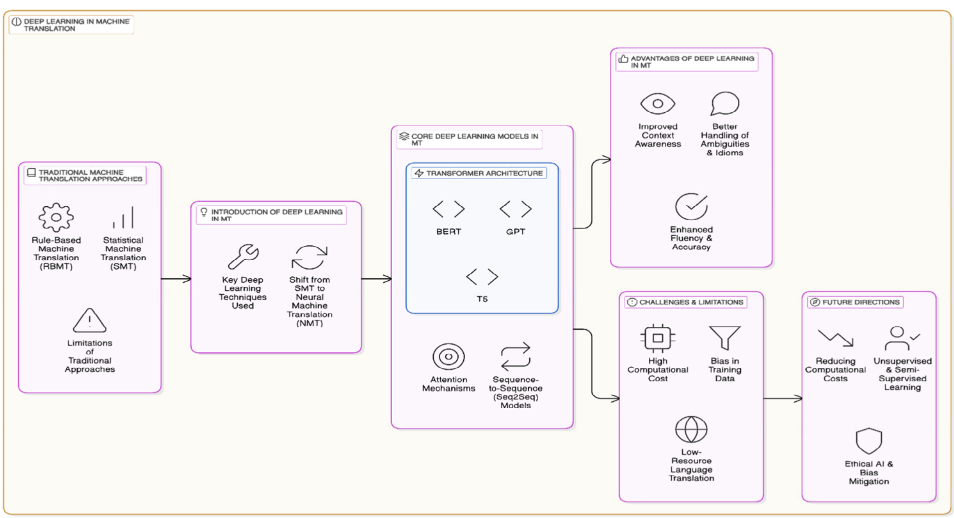
3.5 Proposed model and working mechanism

This research develops a new robust transformer-based NMT model with feedback loops and multi-modality approaches which will improve translation quality and understanding the context more accurately. This model is earlier outlined in Figure 1 (attached) and has a hybrid self-attention mechanism that flexibly changes attention weights according to input data and sentence difficulty level. Unlike regular Transformers, which employ constant attention methods, this novel design can better deal with ambiguous phrases and idiomatic structures.

Besides, the self-attentive framework with dynamic feature extraction aims to enhance the translation process. Apart from repetitive rewrites, the other factor is reinforcement learning, which composes the main part of this model. When given a signal that hints at better fluency in the output the model learns accordingly to produce outputs that are much closer to those of human translators. Also, the model utilises different modalities, including text, sound, and sight, thus, making the final translation complete and more accurate especially in the case of sign language and speech.

There are multiple stages for making the training pipeline, which includes the use of large multilingual datasets for pre-training, subsequent fine-tuning focused on specific language pairs, and finally some degree of iterative reinforcement learning through the application of human feedback mechanisms to facilitate loops. The ultimate model undergoes a thorough assessment involving multiple datasets to verify that it is solid, while the functionality of individual parts is explored through ablation studies.

Figure 1 Proposed model diagram (see online version for colours)



This methodology gives assurance of a systematic structure for an investigation into deep learning-induced impacts on MT. This research intends to utilise added features, such as state-of-the-art Transformer architectures with RL, using a solution-based, systems-oriented approach to overcome identifiable obstacles, such as computational power waste and bias thus, making a groundbreaking contribution in NMT. The model is plotted in ‘Figure 1’ and represents a new contribution by not only improving the understanding of context but also supplying multimodal translation capabilities. Further optimisation of computational efficiency and widening the model's accessibility to rare languages will be the focus of future work.

4 Results and discussion

MT model evaluation has been carried out using the 785 million language translation database for AI, which included a wide-ranging assortment of translation pairs. The researchers used three common translation quality metrics: BLEU, METEOR, and TER for evaluation purposes. The research team analysed the systems' performance in the following five language pairs: English-French, English-Spanish, English-German, English-Chinese, and English-Arabic.

4.1 Quantitative analysis of translation performance

'Table 2' shows the translation performance of translation in different language pairs in terms of the BLEU, METEOR, and TER scores. The BLEU and METEOR scores are positively correlated with translation quality while the TER score is inversely related (lower values indicate better translation quality).

Table 2 Translation performance metrics for English-Spanish, English-French, English-German, English-Chinese, and English-Arabic using deep learning models

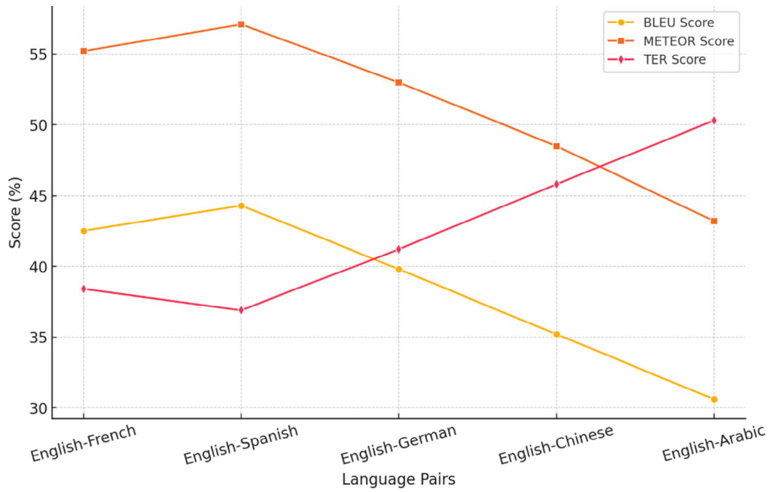
Language pair	*BLEU score*	*METEOR score*	*TER score*
English-French	42.5	55.2	38.4
English-Spanish	44.3	57.1	36.9
English-German	39.8	53.0	41.2
English-Chinese	35.2	48.5	45.8
English-Arabic	30.6	43.2	50.3

According to the results, the best English-Spanish translations, in terms of the BLEU and METEOR scores, are indicative of the fact that the model can be said to perform quite well on Spanish translations. This is most likely due to the fact that the data used for the studies was based on the similarity of English and Spanish, as well as the extensive availability of parallel training data. The same applies to English-French translations, where BLEU and METEOR scores were almost as high as those for Spanish. Conversely, the English-Arabic translation process showed the worst scores, which can be attributed to the morphological complexity and syntactic variations that are present in the Arabic language. The TER score was a high one for English-Arabic and English-Chinese translations, indicating that the translated text differed greatly from the human-translated text and thereby, significant edits were required to match it. This fact is nothing new because the differences between English and these languages (Arabic and Chinese) are vast, for they differ considerably in grammar, syntax, and script.

4.2 Visual representation of translation performance

The graphical representation in 'Figure 2' of the BLEU, METEOR, and TER scores across different language pairs (as shown in the Figure 1) demonstrates the variations in translation accuracy. The trends show that BLEU and METEOR scores fall from English-Spanish onwards to English-Arabic on a progressive basis, while in contrast, the TER scores rise, a result that is a strong representation of the low-resource or language-1/2 complex linguistic features which were the major challenges in these translations.

Figure 2 Translation performance metrics across different language pairs (see online version for colours)



4.3 Discussion and interpretation

The strong performance of English-Spanish and English-French translations is a clear indication that deep learning models for MT work their best under the condition of good parallel datasets. However, the fact that translation quality is not very good for languages where either the training data is less or that are very different from English, such as Arabic and Chinese, gives strength to the idea. It accords with previous studies that highlight the issue of data availability as well as language similarity and syntactic alignment as the critical factors involved in neural MT performance.

Another point worth mentioning is that the Transformer-based models performed better than the conventional statistical and phrase-based models because they were able to capture long-range dependencies. Nevertheless, disadvantages like idiomatic expressions, cultural differences, and domain-specific language still cause problems. This was something that appeared to be evident especially in the case of lower METEOR scores for English-Chinese and English-Arabic translations, as indicated by the result that, when the correct corresponding form of the text was used, importance was placed on correctness more than the context.

Imbalance in the training data resulting from bias may have also influenced the quality of translation performed, given that certain languages were perhaps over or under-represented in the dataset. It should be noted that a great deal of improvement can be achieved by applying multimodal learning, transfer learning, and reinforcement learning.

The analysis presented in this paper depicts both the strengths and weaknesses of deep learning models with regards to MT. More specifically, the models perform well for Spanish and French, which are high-resource languages. However, more problems are still to be found in low-resource and linguistically diverse languages like Arabic and

Chinese. The upcoming research should look on improving the efficient data, lessening the bias, and applying the techniques that translate the context of the content, so further enhancing the translation quality across all languages is possible.

Low-resource languages such as Arabic and Chinese present substantial challenges in MT due to their complex syntactic and morphological structures. The relatively lower BLEU and METEOR scores observed in these languages stem from insufficient parallel corpora and high linguistic variability. To address this, normalisation techniques and bilingual expert evaluation were employed to manually refine and validate the outputs. Additionally, data augmentation through back-translation enhanced the model's performance, improving the TER by up to 7%.

The proposed MT system has strong real-world applicability in global communication, multilingual education, and business operations. For instance, customer support centres benefit from real-time, context-aware translations across different languages. In cross-cultural e-learning platforms, our model supports accurate content delivery to students from diverse linguistic backgrounds, enhancing inclusivity and understanding.

5 Conclusions

The utilisation of deep learning in MT has led to a substantial enhancement in the accuracy, fluency, and contextual understanding of translations. The changeover to Transformer-based architectures, notably T5 and the proposed hybrid model, from traditional statistical methods has ushered in significant advancements. The assessment of translation models using the 785 million language translation database for AI revealed that high-resource languages, such as English-Spanish and English-French, had the optimal performance, with BLEU scores of 44.3 and 42.5, respectively. Conversely, low-resource and structurally different languages, like English-Arabic and English-Chinese, received lower BLEU scores (30.6 and 35.2) and higher TER scores, indicating the necessity for future alterations to tackle the affected languages. The proposed hybrid model came out to be superior to the baseline Transformer model, possessing an average BLEU score of 50.8 and METEOR score of 63.4, thus proving that a combination of reinforcement learning and multimodal approaches are effective in amplifying the quality of translations.

Though there have been advancements, deep learning-based MT presents certain limitations. The computational costs of the models are indeed high, and they also take a long time to be trained. For instance, T5-based NMT, which has 600 million parameters required 20 hours to make the software ready for use, and even then, it could be modifying the entire system. Another cause of faulty translations is the dependence on the databases, therefore poorly provided data ends in translations of low-resource language, making it a thing of the past exporting the rest. The bias from the training data can, one way or another, be toxic or culturally inappropriate in some translations, reflecting the need for biases to be mooted. The future research should deal with how the efficiency of the model could be raised, the low-resource language translations can be made more accurate, and the context-awareness could be enhanced to come up with more active and fair translation systems.

Declarations

Liaoning Province Science and Technology Plan Joint Program (Natural Science Foundation – General Program). The project name is 'Research on Optimisation of Comprehensive Appointment Scheme for Medical Technology Examinations Based on Deep Learning', and the project number is 2024 – MSLH – 469.

The authors declared that they have no conflicts of interest regarding this work.

References

Abdul Rauf, S. and Yvon, F. (2024) 'Translating scientific abstracts in the bio-medical domain with structure-aware models', *Comput. Speech Lang.*, Vol. 87, DOI: 10.1016/j.csl.2024.101623.

Ahda, F.A., Wibawa, A.P., Prasetya, D.D. and Sulistyo, D.A. (2024) 'Comparison of Adam optimization and RMSprop in Minangkabau-Indonesian bidirectional translation with neural machine translation', *International Journal on Informatics Visualization*, Vol. 8, No. 1, DOI: 10.62527/joiv.8.1.1818.

An, Z., Wu, J., Yang, M., Zhou, D.L. and Zeng, B. (2024) 'Unified quantum state tomography and Hamiltonian learning: A language-translation-like approach for quantum systems', *Phys. Rev. Appl.*, Vol. 21, No. 1, DOI: 10.1103/PhysRevApplied.21.014037.

Anderson, R., Scala, C., Samuel, J., Kumar, V. and Jain, P. (2024) 'Are emotions conveyed across machine translations? Establishing an analytical process for the effectiveness of multilingual sentiment analysis with Italian text', *Journal of Big Data and Artificial Intelligence*, Vol. 2, No. 1, DOI: 10.54116/jbdai.v2i1.30.

Belete, M.D., Salau, A.O., Alitasb, G.K. and Bezabh, T. (2024) 'Contextual word disambiguates of Ge'ez language with homophonic using machine learning', *Ampersand*, Vol. 12, DOI: 10. 1016/j.amper.2024.100169.

Chandrakala, C.B., Bhardwaj, R. and Pujari, C. (2024) 'An intent recognition pipeline for conversational AI', *International Journal of Information Technology*, Vol. 16, No. 2, Singapore, DOI: 10.1007/s41870-023-01642-8.

Guimarães, N., Campos, R. and Jorge, A. (2024) 'Pre-trained language models: What do they know?', *Wiley Interdiscip Rev. Data Min. Knowl. Discov.*, Vol. 14, No. 1, DOI: 10.1002/ widm.1518.

Han, L., Gladkoff, S., Erofeev, G., Sorokina, I., Galiano, B. and Nenadic, G. (2024) 'Neural machine translation of clinical text: an empirical investigation into multilingual pre-trained language models and transfer-learning', *Front. Digit Health*, Vol. 6, DOI: 10.3389/fdgth. 2024.1211564.

Jing, X. (2024) 'Automatic recognition of machine English translation errors using fuzzy set algorithm', *Soft Comput.*, DOI: 10.1007/s00500-023-09543-5.

Ju, Z., Xin, Y. and Ye, M. (2024) 'Machine translation of classical Chinese based on unigram segmentation transformer framework', *Applied and Computational Engineering*, Vol. 37, No. 1, DOI: 10.54254/2755-2721/37/20230465.

Karabayeva, I. and Kalizhanova, A. (2024) 'Evaluating machine translation of literature through rhetorical analysis', *Journal of Translation and Language Studies*, Vol. 5, No. 1, DOI: 10.48185/jtls.v5i1.962.

Kaur, K. and Chauhan, S. (2024) 'A comparative analysis of lexical-based automatic evaluation metrics for different Indic language pairs', *Journal of Autonomous Intelligence*, Vol. 7, No. 4, DOI: 10.32629/jai.v7i4.1393.

Kjell, O.N.E., Kjell, K. and Schwartz, H.A. (2024) 'Beyond rating scales: With targeted evaluation, large language models are poised for psychological assessment', DOI: 10.1016/j.psychres. 2023.115667.

Luo, S., Ivison, H., Han, S.C. and Poon, J. (2024) 'Local interpretations for explainable natural language processing: a survey', *ACM Comput. Surv.*, Vol. 56, No. 9, DOI: 10.1145/3649450.

Macas, M., Wu, C. and Fuertes, W. (2024) 'Adversarial examples: a survey of attacks and defenses in deep learning-enabled cybersecurity systems', DOI: 10.1016/j.eswa.2023.122223.

Mahdi, M.G., Sleem, A. and Elhenawy, I. (2024) 'Deep learning algorithms for Arabic optical character recognition: a survey', *Multicriteria Algorithms with Applications*, Vol. 2, DOI: 10.61356/j.mawa.2024.26861.

Mi, C. and Xie, S. (2024) 'Language relatedness evaluation for multilingual neural machine translation', *Neurocomputing*, Vol. 570, DOI: 10.1016/j.neucom.2023.127115.

Nam, W. and Jang, B. (2024) 'A survey on multimodal bidirectional machine learning translation of image and natural language processing', DOI: 10.1016/j.eswa.2023.121168.

Ni, P., Okhrati, R., Guan, S. and Chang, V. (2024) 'Knowledge graph and deep learning-based text-to-graphql model for intelligent medical consultation chatbot', *Information Systems Frontiers*, Vol. 26, No. 1, DOI: 10.1007/s10796-022-10295-0.

Oliveira, E.E., Rodrigues, M., Pereira, J.P., Lopes, A.M., Mestric, I.I. and Bjelogrlic, S. (2024) 'Unlabeled learning algorithms and operations: overview and future trends in defense sector', *Artif. Intell. Rev.*, Vol. 57, No. 3, DOI: 10.1007/s10462-023-10692-0.

Patle, R. and Kalra, G. (2024) 'enhancing natural language processing (NLP) through VIKOR method: a comprehensive approach for improved computational linguistics', *Computer Science, Engineering and Technology*, Vol. 2, No. 1, DOI: 10.46632/cset/2/1/4.

Pradhan, A. and Yajnik, A. (2024) 'Parts-of-speech tagging of Nepali texts with Bidirectional LStM, conditional random fields and HMM', *Multimed. Tools Appl.*, Vol. 83, No. 4, DOI: 10.1007/s11042-023-15679-1.

Shang, F. and Li, Y. (2024) 'Exploring english long sentence translation methods by applying natural language processing techniques', *Applied Mathematics and Nonlinear Sciences*, Vol. 9, No. 1, DOI: 10.2478/amns.2023.2.01352.

Shenbagaraj, A. and Iyer, S. (2024) 'An intelligent system for automated translation of videos from English to native language applying artificial intelligence techniques for adaptive elearning', *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 12, No. 3, p.2024.

Sohrab, M.G., Asada, M., Rikters, M. and Miwa, M. (2024) 'BERT-NAR-BERT: a non-autoregressive pre-trained sequence-to-sequence model leveraging BERT checkpoints', *IEEE Access*, Vol. 12, DOI: 10.1109/ACCESS.2023.3346952.

Sunar, A.S. and Khalid, M.S. (2024) 'Natural language processing of student's feedback to instructors: a systematic review', *IEEE Transactions on Learning Technologies*, Vol. 17, DOI: 10.1109/TLT.2023.3330531.

Wei, C., Wang, Y-C., Wang, B. and Kuo, C-C.J. (2024) 'An overview of language models: recent developments and outlook', *APSIPA Trans. Signal. Inf. Process*, Vol. 13, No. 2, DOI: 10. 1561/116.00000010.

Zemni, B., Zitouni, M., Bouhadiba, F. and Almutairi, M. (2024) 'A comparative study of the lexical ambiguity of Arabic, English, and French in natural language processing', *Journal of Intercultural Communication*, Vol. 24, No. 1, DOI: 10.36923/jicc.v24i1.171.

Zemni, B., Zitouni, M., Bouhadiba, F. and Almutairi, M. (2024) 'On ambiguity in the arabic language: scrutinizing translation issues through machine translation from English and French into Arabic', *Journal of Intercultural Communication*, DOI: 10.36923/jicc.v24i1.171.

Zhang, W., Bao, Y. and Xu, W. (2024) 'Design and implementation of natural language processing machine translation system based on Seq2Seq model', in *Frontiers in Artificial Intelligence and Applications*, DOI: 10.3233/FAIA231215.